

Open Research Online

The Open University's repository of research publications and other research outputs

OpenEssayist: an automated feedback system that supports university students as they write summative essays

Conference or Workshop Item

How to cite:

Whitelock, Denise; Field, Debora; Pulman, Stephen; Richardson, John T. E. and Van Labeke, Nicolas (2013). OpenEssayist: an automated feedback system that supports university students as they write summative essays. In: The 1st International Conference on Open Learning: Role, Challenges and Aspirations.

For guidance on citations see [FAQs](#).

© 2013 The Authors

Version: Version of Record

Link(s) to article on publisher's website:

https://www.arabou.edu.kw/images/stories/files/aouconferenceabstract/denise%20_abs.pdf

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

OpenEssayist: An Automated Feedback System that Supports University Students as they Write Summative Essays

Denise Whitelock¹, Nicolas Van Labeke¹, Debora Field², Stephen Pulman²,
John T. E. Richardson¹

¹Institute of Educational Technology
The Open University
Walton Hall, Milton Keynes, MK7 6AA, UK
Denise.Whitelock@open.ac.uk
John.T.E.Richardson@open.ac.uk
Nicolas.VanLabeke@open.ac.uk

²Department of Computer Science
University of Oxford
Wolfson Building, Parks Road, Oxford OX1 3QD
stephen.pulman@cs.ox.ac.uk
debora@cs.ox.ac.uk

Abstract

OpenEssayist is an automated, interactive feedback system designed to provide an acceptable level of support for students as they write essays for summative assessment. There are two main components to the system: (1) a linguistic analysis engine and (2) a web application that generates feedback for students. The main pedagogical challenge in the e-assessment of free text is how to provide meaningful “advice for action” in order to support students writing their summative assessments. We have built a first working version of the system in which we use unsupervised graph-based ranking algorithms (following Mihalcea & Tarau, 2005) to automatically extract key words, phrases and sentences from student essays. We have developed several external representations of these summarisation techniques. For examples, key words and key phrases can be viewed in a word cloud or in a dispersion graph, and they can be explored and organised into groups. Holistic approaches have also been tested using ‘mash ups’ where key words and key sentences are highlighted in context in the essay itself, helping students to investigate the distribution of key words and its potential implications for the clarity of the narrative. This paper will report the findings from our pilot studies of the interactive models associated with the summarisation techniques.

Keywords

Automated feedback; essay writing; summary; external representations

Introduction

OpenEssayist is a web application that has been designed to assist students while writing their essays for summative assessment. ‘Summative assessment’ is defined as “an assessment that is administered at the end of a learning sequence. It is designed to form a judgement about learning that is often reported in terms of grades or scores and is underpinned by a set of quality assurance processes” (Whitelock, 2011, p. 341). Students who are new to writing essays in higher education and who have returned after a break of many years to undertake a Masters Level qualification often experience difficulty in writing their first summative essay and ‘drop out’ of the course (Simpson, 2003). Therefore the goal of this application is to provide students with feedback on their draft essays before they submit them for summative assessment.

OpenEssayist consists of two major components (1) a linguistic engine and (2) a web application that generates feedback for students (Van Labeke et al, 2013). Understanding how to provide appropriate feedback to students to enable them to move forward with their essay writing is the focus of this paper since providing meaningful feedback or “advice for action” (Whitelock, 2011) needed to be user tested before the system went live in September 2013. A round of supervised, observed user testing was therefore organised, having six participants.

This paper reports on how well the participants understood a range of external representations generated by OpenEssayist in order to assist with essay improvements before submission. The findings have informed the selection of the final representations that will be used for students following a postgraduate module entitled “Accessible online learning: Supporting disabled students”. This postgraduate module runs in September 2013 for about 20 weeks and contributes to a Master of Arts

(MA) in Online and Distance Education. All modules, materials and support are delivered online. Students on this module, as is the case for most of the students at the Open University (OU), are part-time, mature students, many of whom have not been in formal education for a long period of time.

The user testing was designed to understand how to provide meaningful representations of the analysis of draft essays by exploring text analysis outputs and visual analysis outputs

OpenEssayist: The Linguistic Engine

Our approach is to extract different kinds of summaries from the student user's essay, and to present them to her in different ways using a variety of external visualisations. The approach is primarily focused on user understanding and self-directed learning, rather than on essay improvement, and it engages the user on matters of content, rather than pointing out failings in grammar, style, and structure. Our approach contrasts strongly with established Automated Essay Scoring and Automated Writing Evaluation systems (including Criterion (Burststein et al, 2003), Pearson's WriteToLearn (based on Landauer's Intelligent Essay Assessor (Landauer et al, 2003) and Summary Street (Franzke and Streeter, 2006)), IntelliMetric (Rudner et al, 2006) and LightSIDE (Mayfield and Rosé, 2013)). Rather than telling the user in detail how to fix any incorrect and poor attributes of her essay, our system encourages the user to reflect on the *content* of her essay. It uses linguistic technologies, graphics, animations, and interactive exercises to enable the user to comprehend the content of her essay more objectively, and to reflect on whether the essay adequately conveys her intended meanings.

Essay Analysis Output

To run OpenEssayist, the test participant pasted her essay into a web form and selected 'analyse'. The system first returned a representation of the essay that looked very similar to the input, but in which some structural components were identified by different colours (introduction, conclusion, discussion, title, heading, etc.).

In a different view, the key words and key phrases of the essay were presented as a list arranged in rank order of 'key-ness' (which can be thought of as importance, or as representativeness of the text as a whole). In other views, key sentences (those most representative of the text as a whole) were presented both inline in the text 'mash up' view, and in a list in order of importance.

In addition to the core summaries of the essay (key words, key phrases and key sentences), additional specialised data structures were made available, including a 'chord diagram' depicting the results of the key sentence analysis, an 'adjacency matrix' depicting the same, visualised networks depicting the key sentence and key word analyses, and the intersection between the essay's key phrases and words found in the assignment question.

The system was developed by experimenting with a corpus of 267 student essays. While developing the system, much effort has also been devoted to observing and measuring a wide variety of essay attributes (Field et al, 2013). We continue look for ways to exploit these results and, ultimately, to devise effective models of feedback informed by them.

The Empirical Study

The user testing adopted a cognitive walkthrough (Lewis et al, 1990; Polson et al, 1992) as a usability inspected method since it focuses on evaluating a design for user learning through exploring the system, i.e. OpenEssayist. We suspect that our students will prefer to learn how to use the system without resorting to a user guide as supported by the findings of Carroll & Rosson (1987) and Fischer (1991). The system is embodied within an incremental approach to learning and so users will learn how to use and work with features that are of benefit to them. In our case, this means producing external representations of the analysis that will easily translate into changes that students will make to their draft essays.

Participants

Six adults from the Institute of Educational Technology volunteered to take part in the empirical study. They all had experience of either writing summative essays for modules at the Open University UK or had designed essay questions for Open University modules.

Table 1: Demographic profile of participants

Pseudonym	Gender	Mean = 34	Experience of OU essays
Stuart	Male	32	Student who has written OU essays
Albert	Male	33	Student who has written OU essays
Gerald	Male	35	Student who has written OU essays
Nora	Female	28	Student who has written OU essays
Lucy	Female	41	Author of OU essay questions
Sarah	Female	32	Author of OU essay questions

The mean time for each testing session was 59.2 minutes. This allowed the observer to probe participants' reactions after they completed using OpenEssayist. Data from each participant was recorded and transcribed and a systematic manual analysis of screen use was carried out.

Findings

Text Analysis of Essay Structure

The first external visualisation the users were presented with illustrated how OpenEssayist had analysed the structure of the essay.

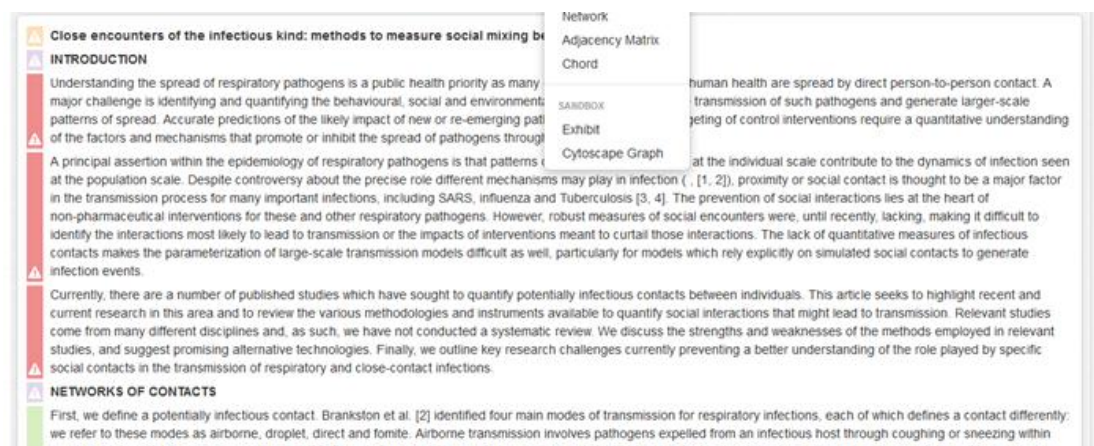


Figure 1: Essay structure prepared by OpenEssayist

The reason for presenting this view was to stimulate user thinking about the structure of their essay. The questions which accompanied this view were as follows:

"Do you think the introduction section as recognised by OpenEssayist is about the right size, or has OpenEssayist got it wrong? Do you think you should try to lengthen the introduction? Or the conclusion?"

Some users commented that displaying the structure of the essay was helpful but that thought that pink was an unsuitable colour for labelling the Introduction. The participants reacted to this pink mark-up as to a red warning sign. They recommended the colour coding needed to be changed. However, with respect to demonstrating the structure of the essay to the user Nora said:

"I can see the benefit because it is talking about the structure. It will help you understand where you need to work in, the different sections, what you are missing maybe you need to fill in a bit more or not."

Visual Analysis of Essay Structure

Another set of representations to illustrate the structure and parts of the assignment was a pie chart depiction and a bullet chart (see Fig. 2). The pie chart illustrated the sizes of the different parts of the

essay as recognised by OpenEssayist whilst the bullet chart illustrated how close the assignment matched the required number of words for the essay.

Lucy remarked on the visual representation in the following way:

“That’s not bad. I can see how much of an introduction it (the essay) has got. How much of a conclusion it has got.”

With respect to the bullet chart, Lucy wanted to know about some ideal values for word counts so that the visualisations could help her change her draft. She said:

“It’s showing me a number of words so my question will be OK I can see that given my story or given the target I have in my TMA, what should it look like? What’s an ideal? What would be expected here?”

This was a constant remark made by the participants that they wanted to see how their draft essay would compare with an “ideal” essay that gained very good marks.



Figure 2: Pie and bullet charts illustrating essay structure

Visual Representation of Essay with Key Words and Key Phrases

OpenEssayist also displays the key words and phrases from the essay. The key word view was an interactive one in which the participants were able to organise the key words into new groups, using as many groups as they wished (see Fig. 3). This interactive task was designed in order to assist the participant in reflecting on the positions and use of particular key terms in the essay

All the participants worked with the clustering task. 3 out of the 6 participants found the task did make them think about how the key words mirrored how the question had been answered but they found the key word and key phrase list more useful, saying that it acted as a summary of the main issues that had been addressed.

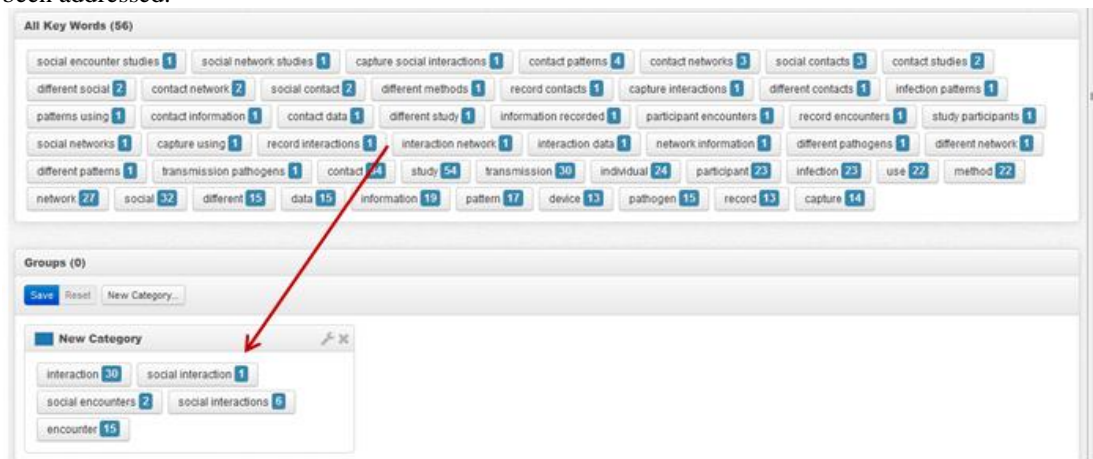


Figure 3: Clustering key words

Summary

OpenEssayist presents the essay's key sentences as a list (see Fig. 4). Stuart remarked that:

“The summary is an extension of the key words and phrases and will make me think about whether this is really what I wanted to say in the essay.”

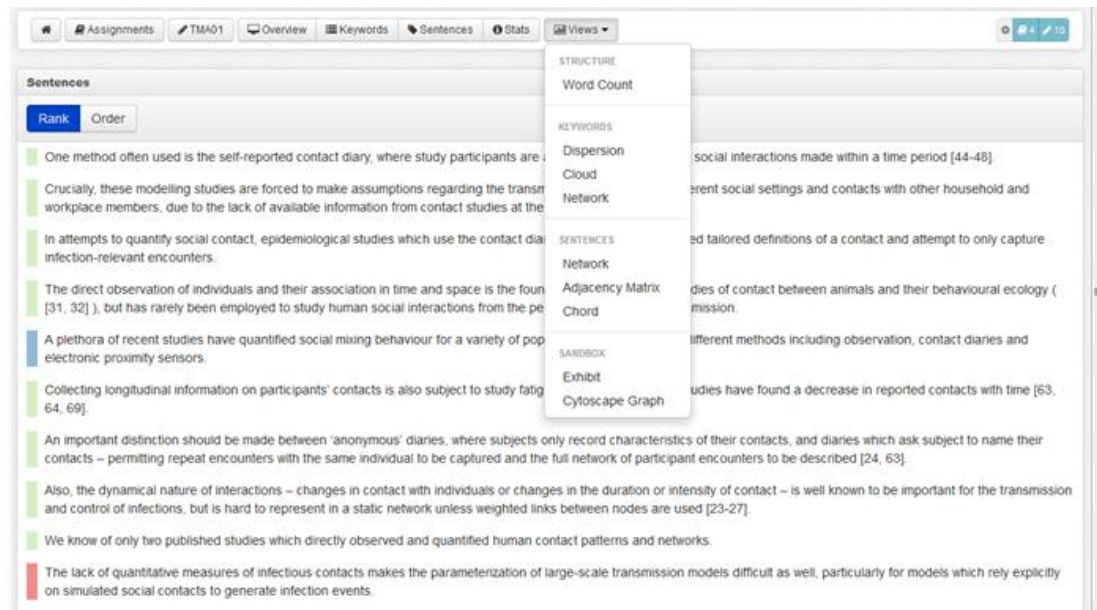


Figure 4: Key sentences summary

Key Words and Phrases Highlighted in the Essay

Fig. 4 illustrates the raw text of the essay with features extracted by the EssayAnalyser in context. In this instance the key words were displayed but key phrases could also be displayed. Sarah commented about this visualisation in the following way:

“So actually now on reflection, now I am looking at it, yes another way of looking to see how well you’ve structured and how well you’ve got the right information in your introduction, conclusion particularly. Personally for me would be good because they were probably one of my weakest areas as a student. So that would be useful.”



Figure 5: Key words illustrated in main text of essay

Complex External Visualisations

The following three visualisations were also tested with the participants. These included the dynamic, moving spring (Fig. 6), the chord diagram (Fig. 7) and the adjacency matrix (Fig. 8). The reasoning behind the chord diagram and the adjacency matrix view was to help the user see how their arguments were being progressed in the essay. The reasoning behind the dynamic, moving spring view was to

give the user some understanding of how the key words were chosen by the system, and how they were related to each other.

These were indeed complex representations which all the participants found difficult to understand. An example of their reactions is illustrated by Stuart's remark:

“A lot of people struggle with visual representation of things. I need a lot of help to intepret that and what that might mean.”

According to Bertin (1981), in order to perceive a graphic, two stages of perception are required. The first is to identify the elements in question and the second is to determine what is the relationship among those elements. The problems that arose for the participants with these complex visualisations were that the given elements were not immediately legible or comprehensible. More importantly, for the participants, the relationship between the signs in the graphic imposed too much of a cognitive load and the participants gave up trying to interpret them. They asked questions about how these visualisations could help the improve their essays and Lucy mentioned:

“If I was a linguist and I was analysing structure of text that might be interesting. I am not sure a student would know what to do with that if I'm perfectly honest.”

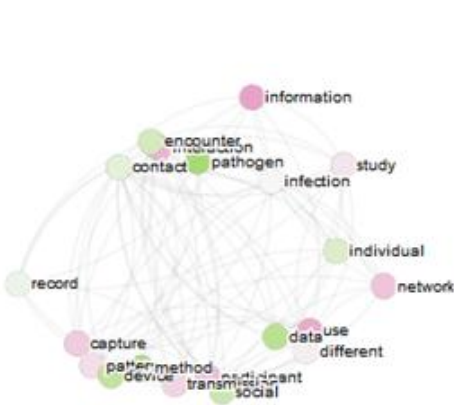


Figure 6: Key word spring

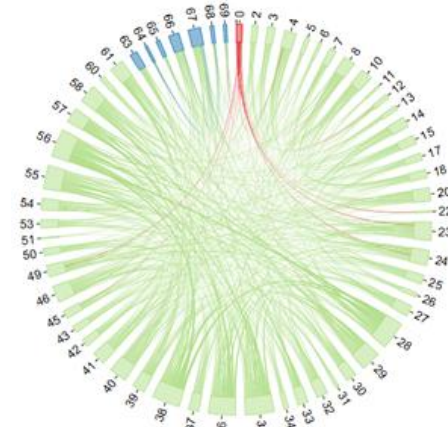


Figure 7: Chord diagram

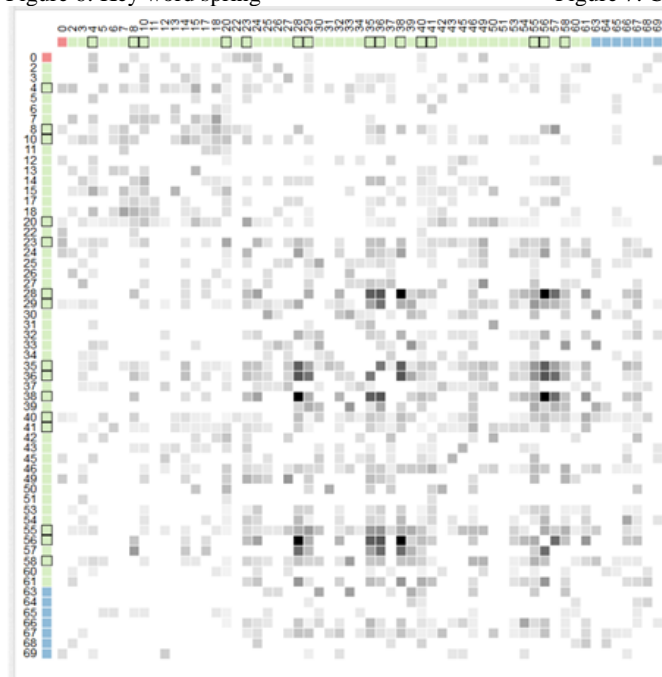


Figure 8: Key sentence adjacency matrix

Conclusion

The user testing set out to understand how meaningful the external representations presented by OpenEssayist were to a group of users familiar with the types of summative assessments used by the Open University. A mixture of texts and graphics were understood by the participants with respect to the structure of the essay, i.e., the introduction, discussion and conclusion. The representations encouraged reflection on how key words and phrases were positioned within the text and whether the draft essays were adequately addressing the assignment question.

However, the more complex visualisations were more difficult for these “naive users” to comprehend. One suggestion from a participant was that these complex graphics could not be acted upon to facilitate further visual inspection. Although some interactivity was available, these did not provide further clarity for the user. These findings support the work of Wise et al (1995) who concluded that:

The success of other text visualisations will likely be determined by whether the user can manipulate them along the lines of their analytical intuitions.

Our future work into visual representations for OpenEssayist will concentrate on the visual metaphors that capture how concepts become understandable and the decisions that can be made for essay improvements.

Another area for exploration is animating the transitions (Yee et al, 2001) from one view of a representation to another that reduces the user’s incomprehension. Understanding graph structure, which is at the core of our representations, is indeed a well known and long term problem for creators of information visualisations. For us the complex graphics in which we were trying to illustrate nearness produced confusion.

These are interesting times for formative assessment. The provision of timely and pertinent formative feedback is becoming more widely available with the rapid developments that are taking place in web technologies. The challenge is how to exploit new insights from the Natural Language Processing domain to improve formative feedback.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (grant numbers EP/J005959/1 and EP/J005231/1).

We would also like to thank Evaghn Desouza for his assistance with the user testing.

References

- Alden, A., Van Labeke, N., Field, D., Pulman, S., Richardson, J.T.E. & Whitelock, D. (2013) Using student experience to inform the design of an automated feedback system for essay answers. *Proceedings of the 2013 International Computer Assisted Assessment (CAA) Conference*, Southampton. Retrieved from http://caaconference.co.uk/wp-content/uploads/Alden_caa2013_submission.pdf
- Bertin, J. (1981). *Graphics and Graphic Information-Processing* (W. J. Berg and P. Scott, Trans.), Berlin: New York: de Gruyter.
- Burstein, J., Chodorow, M. & Leacock, C. (2003). CriterionSM online essay evaluation: An application for automated evaluation of students essays. In J. Riedl and R. Hill (Eds.), *Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence* (pp. 3 – 10). Cambridge, MA: MIT Press.
- Carroll, J.M. & Rosson, M.B. (1987). The paradox of the active user. In J.M. Carroll (Ed.) *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction* (pp. 80-111). Cambridge: Bradford Books/MIT Press.
- Chase, J.A. & Houmanfar, R. (2009). The Differential Effects of Elaborate Feedback and Basic Feedback on Student Performance in a Modified, Personalised System of Instruction Course. *Journal of Behavioural Education*, 18, 245-265.
- Field, D., Richardson, J., Pulman, S., Van Labeke, N. & Whitelock, D. (2013). Reflections on characteristics of university students’ essays through experimentation with domain-independent natural language processing techniques. *Proceedings of the 2013 International Computer Assisted*

- Assessment (CAA) Conference, Southampton. Retrieved from http://caaconference.co.uk/wp-content/uploads/Field_caa2013_submission.pdf
- Fischer, G. (1991). Supporting learning on demand with design environments. *Proceedings of the International Conference on the Learning Sciences*, 1991 (Evanston, IL, August): 165-172. Charlottesville, VA: Association for the Advancement of Computing in Education.
- Franzke, M. & Streeter, L.A. (2006). *Building student summarisation, writing and reading comprehension skills with guided practice and automated feedback*. White paper, Pearson Knowledge Technologies. Accessed: 14 May 2013.
- Kulhavy, R.W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232.
- Kulik, J.A. & Kulik, C.-L.C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58, 79-97.
- Landauer, T.K., Laham, D. & Foltz, P.W. (2003). Automatic essay assessment. *Assessment in Education: Principles, Policy and Practice*, 10(3), 295-308.
- Lewis, C., Polson, P., Wharton, C. & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proceedings ACM CHI'90 Conference* (pp. 235-242). Seattle, WA.
- Mayfield, E. & Rosé, C.P. (2013). LightSIDE: Open source machine learning for text. In M.D. Shermis & J. Burstein (Eds.), *Handbook of Automated Essay Assessment Evaluation* (pp.124-135). Taylor & Francis.
- Mihalcea, R. & Tarau, P. (2004). TextRank: Bringing Order into Text. *Proceedings of Empirical Methods in Natural Language Processing EMNLP 2004*, (pp. 404-411). <http://aclweb.org/anthology/W/W04/#3200?>
- Mislevy, R.J., Behrens, J.T., Bennett, R.E., Demark, S.F., Frezzo, D.C., Levy, R., Robinson, D.H., Rutstein, D.W., Shute, V.J., Stanley, K. & Winters, F.I. (2010). On the Roles of External Knowledge Representations in Assessment Design. *Journal of Technology, Learning and Assessment*, 8(2).
- Nielsen, J. & Mack, R.L. (Eds.) (1994). *Usability Inspection Methods*. New York: John Wiley & Sons.
- Paranyushkin, D. (2012). *Visualisation of Text's Polysingularity Using Network Analysis*. Nodus Labs, Berlin, Germany. Retrieved from <http://www.noduslabs.com/publications/text-polysingularity-network-analysis.pdf>
- Polson, P., Lewis, C., Rieman, J. & Wharton, C. (1992). Cognitive walkthrough: A method for theory-based evaluation of user interface. *International Journal of Man-Machine Studies*, 36, 741-773.
- Preece, J. & Janvier, C. (1992). A Study of the Interpretation of Trends in Multiple Curve Graphs of Ecological Situations. *School Science and Mathematics*, 92(6), 299-306.
- Rudner, L.M., Garcia, V. & Welch, C. (2006). An evaluation of the IntelliMetricsSM essay scoring system. *The Journal of Technology, Learning and Assessment*, 4(4).
- Simpson, O. (2003). *Student Retention in Online, Open and Distance Learning*. London; Sterling, VA: Kogan Page.
- Van Labeke, N., Whitelock, D., Field, D., Pulman, S. & Richardson, J. (2013). *OpenEssayist: Extractive Summarisation & Formative Assessment of Free-Text Essays*. Workshop on Discourse-Centric Learning Analytics, 3rd Conference on Learning Analytics and Knowledge (LAK 2013), Leuven, Belgium
- Whitelock, D. (2011). Activating Assessment for Learning: are we on the way with Web 2.0? In M.J.W. Lee & C. McLoughlin (Eds.), *Web 2.0-Based-E-Learning: Applying Social Informatics for Tertiary Teaching* (pp. 319-342). IGI Global.
- Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A. & Crow, V. (1995). Visualising the Non-Visual: Spatial analysis and interaction with information from text documents. *Proceedings of the IEEE*. Retrieved from https://www.cs.duke.edu/courses/cps296.8/current/papers/vis_non_visual.pdf
- Yee, K.P., Fisher, D., Dhamija, R. & Hearst, M. (2001). *Animated Exploration of Dynamic Graphs with Radial Layout*. Paper presented at IEEE Symposium on Information Visualization, Chicago, USA. Retrieved from [http://www.new.artoolkit.org/courses/InfoVis/papers/InfoVis2001%20Animated%20Exploration%20of%20Dynamic%20Graphs%20with%20Radial%20Layout%20\[Ka-Ping%20Yee\].pdf](http://www.new.artoolkit.org/courses/InfoVis/papers/InfoVis2001%20Animated%20Exploration%20of%20Dynamic%20Graphs%20with%20Radial%20Layout%20[Ka-Ping%20Yee].pdf)